Proceedings of the ASME 2017 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2017
August 6-9, 2017, Cleveland, USA

DETC2017/DTM-XXXXX

IMPLICATIONS OF DATA-DRIVEN MODELS FOR DESIGN THEORY AND METHODOLOGY

Mark Fuge
Department of Mechanical Engineering
University of Maryland
College Park, Maryland 20740
Email: fuge@umd.edu

ABSTRACT

This presentation 1) reviews the outcomes of a one-day symposium on Data-Driven Design held at the University of Maryland on October 16th, 2016; and 2) discusses the implications of those outcomes on new Design Theory and Methodology research. These outcomes include opportunities and challenges for Data-Driven Design, ranging from modeling to verification to establishing researcher culture. This presentation organizes those opportunities and challenges and then poses a series of questions to the DTM community to catalyze discussion about how to prioritize, address, or add to these challenges. My intent is for the presentation to enable group discussion during the conference and gather feedback about key challenges and calls to action for the DTM community. In this respect, I intend for this presentation to raise more questions and discussion than it answers.

INTRODUCTION

The use of Data-Driven models—statistical and computational models that take in data from designs or designers to draw conclusions about design—have become increasingly common in design. This has grown in different forms and names throughout different communities; within ASME IDETC this includes subsets of "Design Informatics" in CIE, "Data-Driven Design" in DAC, and "Design Computing" in DTM. However, identifying what approaches are *Data-Driven* or not is confusing; after all, are not all design decisions ultimately driven by *some* data?

To help identify common and important challenges to applying Data-Driven models for understanding design, the University of Maryland hosted a one-day symposium on Data-Driven Design on October 14th, 2016. This symposium brought together academic, industrial, and government researchers from various fields—Mechanical Engineering, Aerospace Engineering, Manufacturing, Human Factors, Computer Science, and Naval Architecture, among others—to discuss different opportunities and challenges that arise when using Data-Driven models to understand and influence design.[1]

This technical presentation summarizes that discussion and key challenges for the wider ASME community. In particular, the presentation focuses on 1) what does it entail for a given model or analysis to be *Data-Driven* (with illustrative examples) compared to other approaches to studying design, and 2) how and when are Data-Driven models useful (and harmful) to advancing Design Theory and Methodology?

OPPORTUNITIES AND APPLICATIONS FOR DATA-DRIVEN MODELS OF DESIGN

What do we mean when we say a model is "Data-Driven" and for what kinds of problems are such models useful? To understand those questions, the participants discussed several examples, organized by where in a product's life-cycle the *data* in a data-driven model comes from.

[1] Speakers included representatives from: Autodesk, Carnegie Mellon University, Clemson University, Duke University, the Navy, the Department of Energy, the National Institute of Standards, the National Science Foundation, the University of Buffalo, the University of Maryland, and the University of Urbana Champaign; along with participants from many other US and UK institutions.

Conceptual Design The discussed examples used data-driven models to navigate, limit, or provide performance bounds on otherwise intractably large design spaces. This focus arose from the fact that conceptual design tends to be both ill-defined and expansive; exploring and characterizing such spaces is difficult and high dimensional. For example, one can use past designs or simulations to learn useful sub-spaces to aid human exploration, use past simulations to learn fast approximations of physical phenomena (*e.g.*, CFD or topology optimization) to compare conceptual designs, or use external sources (like Wikipedia) to learn useful common knowledge (*e.g.*, about functional relationships between parts) that can then combine with past design methods.

Embodiment Design The discussed examples dealt with representing embodiment performance targets of interest and how best to encode the right properties of a system so that you get good predictive accuracy (in the worst case) or even causal mechanisms (in the best case). For example, using data-driven models like Neural Networks to approximate assembly complexity (and thus design faster-to-assemble systems), learning structured probabilistic graphical models for motion patterns in footwear to modify specific part dimensions or components, or learning patterns in how human make decisions during complex system design tasks to better predict choices.

Manufacturing The discussed examples tackled representation issues across different parts of the *digital thread*, given that they do not really talk to one another. This lack of communication complicates purely physics-based simulations, and thus predictive data-driven models can provide manufacturing design insights. For example, by building predictive models of part manufacturing (*e.g.*, time series power draw on a CNC machine) one can compare the actual and predicted curves to uncover part features that complicate CNC machining, or by using data-driven models to visualize and fuse together different information about a part as it is being made so as to improve the design.

In-Use The discussed examples discussed how to connect real-world product or human behavior to design choices and ensure that data-driven models capture the right user behavior dynamics. Particularly with humans, system dynamics will likely alter in response to design changes. For example, in Internet of Things (IoT) applications, products in use can both influence future designs of products, but also modify their own behavior (*e.g.*, a tractor coordinating via data with a manufacturer to improve reliability and performance). Likewise, autonomous cars or planes can learn from observing human behavior to better design themselves to communicate system intent.

CHALLENGES TO DATA-DRIVEN MODELS AND IMPLICATIONS FOR DTM RESEARCH

The participants identified several key challenges and questions that could affect how data-driven methods could contribute to Design Theory and Methodology research and practice.

Verifying and Validating Data-Driven Design Models A central issue in data-driven models is verifying the conditions under which they perform well. This problem rears its head in design in a couple of ways: 1) What are the outcome measures we care about? Are traditional accuracy or recall measures sufficient, or must we explicitly map them to design outcomes from industry? 2) Are there really any "common" design problems, or are all problems so different and dynamic that models should not predict based on past data from different problems? 3) How do we do repeatable or reproducible experiments when humans (as users or designers) are in the loop? How do we build "Trust" into data-driven models? Can we complement data-driven models with traditional physics-based models to test validity?

Generalization and Extrapolation of Data-Driven Models To what extent are Data-Driven Models "transferable" between design problems? How do we handle systems we have not yet seen? To what extent is past data really predictive of future systems, given that we (both designers and humans) learn from past systems? Could we even change our own interaction with existing systems over time? To what extent can data-driven design models be "explainable?" Can we bound the problems that data-driven design is ill-suited to solve?

Standards, Benchmarks, and Practice for Data Collection, Sharing, and Training How do we capture, process, store, and share data (size, speed, formats, *etc.*)? What if the data we get (*e.g.*, accelerometer data) looks far removed from the actual design system of interest (*e.g.*, the manufacturability of a feature)? How do we ensure that we have gathered sufficient data for the task at hand? How do we measure progress in the field toward concrete goals or benchmarks?

Regulatory, Ethical, and Intellectual Property Concerns How do we regulate systems controlled or designed by data-driven models? Could data-driven models inform regulation itself? How do we ensure our models do not learn harmful and difficult to detect bias directly into the products we use? How do we debug such behavior? How do we interpret the ASME ethics code in the context of data-driven models? How will we address privacy and security concerns, if external data can influence physical products? How do we incentivize a research culture amendable to data-driven design research, for example, sharing data or benchmarks?

Understanding Fundamental Observability, Complexity, and Causality Limits for Data-Driven Models How do we know when we can detect emergent behavior within complex systems or designs? When is enough data sufficient? How do we balance causation with prediction in data-driven models?

ACKNOWLEDGMENT

We acknowledge support from the Center for Engineering Concepts Development and the Mechanical Engineering department at the University of Maryland through a Minta Martin grant.